图书情报工作
第62卷 第3期 2018年2月

美国开放政府数据元数据标准及启示*

■ 司莉[1] 赵洁[2]

[1] 武汉大学信息资源研究中心 武汉 430072 [2] 武汉大学信息管理学院 武汉 430072

摘要:[目的/意义]以美国开放政府数据网站 Data. gov 中的元数据标准为例,分析其元数据体系及具体标准,以期为我国开放政府数据元数据标准的构建提供参考。[方法/过程]采用实例分析的方法,归纳总结美国开放政府数据元数据标准的体系结构。[结果/结论]美国开放政府数据元数据标准分为数据集内容与数据集格式描述元数据标准,并针对原始数据集与地理空间数据集采用不同标准描述;并指出我国在构建自身开放政府数据元数据标准时可借鉴 Data. gov 中的元数据标准体系。

关键词:元数据 开放政府数据 元数据标准 Data. gov

分类号:G254.31

DOI:10.13266/j. issn. 0252 -3116. 2018. 03. 011

开放政府数据是在数据开放运动和政府信息公开的驱动下产生的一类数据。政府数据是指由政府或政府所属机构产生或委托产生的数据与信息,开放即可以被任何人免费使用、重用和再传播[1]。由于政府数据开放可以增加政府工作透明性、增强数据的社会和商业价值、提高公民对政府活动的参与度[1],世界各国包括英国、美国、澳大利亚、加拿大以及中国等国家积极开展开放政府数据活动,并构建各自的开放政府数据平台,以提供开放政府数据的统一获取与利用。

开放政府数据平台通常以数据目录的形式集成多源数据集,元数据标准是数据目录管理数据集的重要方式,已被英国、中国等国家列为开放政府数据质量的考察指标之一[2-3]。我国国家开放政府数据平台正处于构建之中,现有的开放政府数据平台多为地方政府数据平台,如北京、上海等地方开放政府数据平台,而各个地方平台元数据标准存在标准不统一[4]、元数据信息匮乏[4]、数据集描述不全面[5]、缺乏可机读格式[5]、互操作水平低[6]等问题。基于元数据标准对开放政府数据平台中数据集整合的重要性,及我国当前开放政府数据元数据标准现状,亟需对开放政府数据元数据标准的构建展开研究。国内学者目前已针对开放政府数据元数据标准展开相关研究,多以国外成熟的开放政府数据平台为研究对象,如英国[7]、澳大利亚[8]、加拿大[9]、新西兰[10]等国家的开放政府数据元数据标准。赵蓉英、梁志森、段培培[7]以英国开放政府数据平台 Data. gov. uk 为研究对象,对其所使用的 CKAN 格式记录(包括 CSV 和 JSON)和 GEMINI 地理空间元数据标准,从文件结构、元素组成及规则等方面总结其特点。黄如花与李楠[8]选取澳大利亚开放政府数据平台 Data. gov. au 为研究对象,对其所使用的 3 种元数据标准即 AGLS 元数据标准、ANZLIC 地理空间元数据标准和数据目录词表 DCAT,从各标准的元素组成、数据格式、语法结构等方面展开了调查与分析。王立清与唐宇萍[10]则选取了澳大利亚 AGLS 和新西兰 NZGLS 元数据标准进行了研究,主要包括标准的建立基本情况、元数据元素定义、相关限定词等。于梦月、翟军、林岩[5-6]对 W3C 的正式推荐标准 DCAT 和美国纽约州的元数据方案进行了介绍,并分析和总结了美国、欧盟和爱尔兰政府开放数据元数据建设的成果和特点。武琳与黄颖茹[9]则在详细梳理美、英、加和欧盟的相关元数据政策和标准的基础上,对各个国家元数据标准的元数据格式、元数据框架、元素、数据目录词表、受控词表等方面进行比较分析。黄如花与林焱[11]对开放政府数据水平较高的英国、美国、加拿大、新西兰、欧盟的政

* 本文系国家自然科学基金项目"大数据环境下科研数据机构库联盟形成机理及其服务研究:以'985'高校为实证对象"(项目编号:71573198)研究成果之一。

作者简介:司莉(ORCID:0000 -0003 -1028 -8338),教授,博士生导师;赵洁(ORCID:0000 -0002 -6578 -1413),博士研究生,通讯作者,E-mail:zhaojie_shuique@163.com。

收稿日期:2017 -07 -24 修回日期:2017 -11 -20 本文起止页码:86 -93 本文责任编辑:徐健

府数据开放门户及其相关公共部门的元数据描述规范进行了调研。而在上述综述中,并未见到专门针对美国开放政府数据元数据标准的详尽研究。

美国是实施开放政府数据政策较早的国家,其开放政府数据平台 Data. gov 于 2009 年 5 月构建,截至 2017 年 11 月 15 日,已有 198 个组织的 197 990 个数据集在该平台公开发布,平台中的数据集仍在不断更新。Data. gov 平台以数据目录的形式组织并管理各个机构拥有的数据集,即各个机构的数据集并不直接存储于平台中,而是在采用统一元数据标准描述后,基于元数据收割的方法,将数据集的基础信息提取至数据目录中,集中呈现数据集的概要信息,并提供数据集的获取链接和下载链接。Data. gov 平台定期重新收割机构拥有的数据集的元数据信息,以及时更新数据目录。本文选择美国开放政府数据 Data. gov 作为研究对象,分析其元数据体系和标准,旨在补充开放政府数据元数据标准方面的研究,为我国开放政府数据元数据标准的构建提供借鉴。

1 美国开放政府数据元数据标准

Data. gov 是以数据目录的形式集中呈现数据集信息,该平台并不直接物理存储数据资源,而是采用元数据收割的方式获取数据集的目录信息,通过目录信息链接到具体的数据集。元数据收割的实现基于数据集的统一元数据标准描述。Data. gov 按照原始数据集、地理空间数据集和数据工具 3 个门类组织平台中开放的数据资源,并分别采用不同的元数据标准对这 3 类具有不同特征的数据资源进行描述。数据工具是指基于原始数据集和地理空间数据集开发的各种 API。本文主要分析描述原始数据集与地理空间数据集的元数据标准。笔者根据对数据集描述方面的不同,将涉及到的元数据标准划分为数据集内容描述元数据标准与数据集格式描述元数据标准两大类,如表 1 所示:

表 1 美国开放政府数据平台 Data. gov 元数据标准

Data. gov 元数据标准类型	元数据标准	描述对象
数据集内容描述元数据标准	项目开放数据元数据标准(*Project Open Data metadata schema*,POD v1.1)	原始数据集
	ISO 19115-2 图像和栅格数据元数据标准	地理空间数据集
	数字地理空间元数据内容标准(*Content Standard for Digital Geospatial Metadata*,CSDGM)	
数据集格式描述元数据标准	JSON	使用"POD v1.1"描述的数据集
	ISO 19139 地理信息-元数据-XML 标准实现(*ISO 19139 Geographic Information-Metadata-XML schema implementation*,*ISO 19139*)	使用"ISO 19115-2 图像和栅格数据元数据标准"描述的数据集

1.1 数据集内容描述元数据标准

该标准是指对数据集自身信息进行描述的元数据标准。不同类型的数据集具有不同的内容描述方面,所采用的元数据标准也不相同。根据 Data. gov 中的资源类型,将数据集内容描述元数据标准分为原始数据集内容描述元数据标准与地理空间数据集内容描述元数据标准。

1.1.1 原始数据集内容描述元数据标准 原始数据集由联邦政府或其所属机构提供。Data. gov 使用统一的开放政府数据元数据标准——*Project Open Data metadata schema*[12](以下简称 POD v1.1)来描述原始数据集信息。该标准是开放数据项目(Project Open Data)提出的数据集应遵循的元数据标准,它是基于数据目录词表[13](*Data Catalog Vocabulary*,DCAT)构建的专用于数据集描述的层级词表。2015 年 2 月已更新至第 2 版即 1.1 版本。以下分别从描述内容、字段类别、元数据元素 3 个方面对其进行分析。

(1)描述内容。描述内容主要涉及数据集内容与元数据标准两个方面。内容描述信息主要包括内外部内容信息的描述:外部内容信息,即数据集从创建到发布整个过程中所涉及到的人、机构、时间、空间、许可、权利等信息;内部内容信息,如标题、描述、标签等信息。元数据标准描述信息是指元数据标准版本等信息。

(2)字段类别。按照元数据字段描述对象的不同,该标准中的字段分为目录字段(catalog fields)、数据集字段(dataset fields)、数据集发布字段(dataset distribution fields)三种类型。元数据字段可划分为必要字段(required fields)与非必要字段(non-required fields)。必要字段又分为不可变必要字段(required fields, always)和可变必要字段(required-if(conditionally required))。不可变必要字段即每个数据集中必须含有该字段,可变必要字段即在数据集满足某种条件的情况下必须包含该字段,具体条件依元素不同而不同,如机构代码"局代码(bureauCode)"这一字段只有在数据集所属机构为美国联邦政府时才为必要字段,

“发布(distribution)”字段只有当数据集有“accessURL”或“downloadURL”时才是必要字段。扩展字段(expanded fields)为非必要字段。

(3)元数据元素。POD v1.1 元数据标准字段划分及相应元素见表 2。由该表可知,POD v1.1 元数据标准共含有 45 个元素,涉及到数据目录、数据集、数据集发布 3 个方面的数据描述。

表 2 POD v1.1 元数据标准字段划分及包含元素

字段划分	包含元素
目录字段(catalog fields)	不可变必要字段:元数据标准版本、数据集; 扩展字段:元数据背景、元数据目录标识、元数据类型、数据字典。
数据集字段(dataset fields)	不可变必要字段:标题、描述、标签、最新更新时间、发布者、联系人姓名和邮箱、唯一标识符、公开获取等级、局代码、项目代码; 可变必要字段:许可、权利、空间信息、时间信息、发布; 扩展字段:元数据类型、数据集发布频率、数据标准、数据质量、数据字典、数据字典类型、数据集计划、发布日期、数据集语言、数据集登录页面、信息技术唯一投资标识符、相关文档、记录系统、主题类别。
数据集发布字段(dataset distribution fields)	可变必要字段:获取 URL、下载 URL、媒体类型 扩展字段:元数据类型、数据标准、数据字典、数据字典类型、用户可读描述、格式、标题。

①目录字段:用来描述整个公共数据列表目录文件,数据目录是多个数据集描述的集合。目录字段共包含 6 个元素,其中 2 个元素属于不可变必要字段、4 个元素属于扩展字段。不可变必要字段描述了数据目录包含的具体数据集对象及数据集遵循的元数据标准版本;扩展字段则是对描述数据集的元数据标准的描述。

②数据集字段:用来描述单个数据集对象的内容特征,共有 29 个元素,其中不可变必要字段、可变必要字段、扩展字段分别包含 10 个、5 个、14 个元素。从数据集的具体字段进行分析,不可变必要字段主要描述与数据集创建、主题、识别、公开、归属有关的基本信息;可变必要字段则主要描述数据集的发布信息、时间空间信息、许可权利信息等;扩展字段是对数据集元数据信息、数据质量、数据更新情况等进行描述。

③数据集发布字段:用来描述与数据集获取有关的信息,共有 10 个元素,其中,无不可变必要字段,可变必要字段与扩展字段分别有 3 个、7 个元素。虽无不可变必要字段,但要求每一组数据集发布字段包含一个获取 URL 或下载 URL。可变必要字段主要为数据集的具体获取地址,包括获取地址和下载地址。扩展字段则为对数据集的元数据信息的描述。

根据上述分析可知,目录字段、数据集字段、数据集发布字段是层层嵌套的关系。目录字段中的“数据集(dataset)”子元素可细分为数据集字段中的所有元素,数据集字段中的“发布(distribution)”子元素可细分为数据集发布字段中的所有元素。

在设计元数据标准时,Data.gov 是结合该平台对数据集的呈现方式设计的,先以目录形式呈现所有数据集的摘要列表,具体包含数据集标题、所属机构、数据集内容摘要、数据集获取格式等信息;再对单个数据集进行内容详细描述,并对有数据获取和下载方式的数据集进行数据发布信息的描述,从而帮助用户找到所需要的数据集。同时,目录字段、数据集字段、数据集发布字段均以必要字段(包括不可变与可变必要字段)和非必要字段(拓展字段)为框架进行具体属性的设计。不可变必要字段为各类型字段的核心部分,描述了代表该类型字段的核心属性;可变必要字段针对特定包含相关字段信息的数据集;扩展字段则主要是对元数据信息、所使用的数据字典及格式、数据集的呈现、数据集的其他相关信息进行声明。

1.1.2 地理空间数据集内容描述元数据标准 地理空间元数据用来描述地图、地理信息系统文件、图像及其他基于位置的数据资源[14],通常被包含在 Data.gov 的 GeoPlatform.gov 子平台中。Data.gov 中的地理空间数据集为 197 990 个(截至 2017 年 11 月 15 日的统计),约是非地理空间数据集的 1.8 倍。地理空间数据集的描述,除需遵循 POD v1.1 元数据标准外,还要遵循专门的地理空间元数据标准。地理空间数据集包含的地理信息主要包括 3 个类型:一是记录地理实体的空间数值和属性特征,如经纬度、比例尺等元素;二是标记空间数据所依附载体的信息,如所参考的地理信息系统、地图显示框等元素;三是针对资源对象或用户的元素,如证明对象来源和约束用户使用资源的元素等。Data.gov 中的地理空间数据主要遵循两种地理空间元数据标准,分别为 ISO 19115-2 图像和栅格数据元数据标准以及数字地理空间元数据标准(*Content Standard for Digital Geospatial Metadata*,CSDGM)。

(1)ISO 19115-2 图像和栅格数据元数据标准。ISO 19115 是地理信息的内容和描述标准,Data.gov 使

司莉，赵洁. 美国开放政府数据元数据标准及启示[J]. 图书情报工作,2018,62(3):86－93.

用该标准的第二部分：图像和栅格数据（ISO 19115－2 Geographic Information-Metadata-Part 2：Extensions for Imagery and Gridded Data，目前版本为2009年版）进行描述，如表3所示：

表3　ISO 19115－2 图像和栅格数据元数据标准主要部分

主要部分	描述内容	描述对象
元数据根信息	包含元数据信息的根元素	元数据
空间表示信息	地理空间表示信息	数据集
参考系统信息	空间和时间参考系统信息	数据集
元数据扩展信息	描述资源的用户特定扩展信息	元数据
识别信息	唯一识别资源的信息	数据集
内容信息	资源包含的物理参数和其他属性信息	数据集
发布信息	关于资源发布者和如何获取资源的信息	数据集
数据质量信息	资源的质量、处理步骤和来源信息	数据集
描绘目录信息	资源使用的识别描绘目录的信息	数据集
元数据限制信息	元数据和资源的使用限制信息	元数据
应用标准信息	构建数据集的应用标准信息	数据集
元数据维护信息	元数据和其所描述资源的维护信息	元数据
获取信息	数据获取的工具、平台、操作信息和其他相关信息	数据集

表3列出了该元数据标准包含的13个主要部分，并对这13个主要部分所描述的内容进行了说明。由表3可知，ISO 19115－2元数据标准包含了与数据集有关的初期的内容描述、中期的质量评估与数据发布、后期的数据使用与维护的各个方面，完整性较好。根据描述对象的不同，可以将元数据标准中的这13个方面分为对数据集与对元数据标准的描述。其中，数据集具体地理特征主要通过空间表示信息、参考系统信息中的字段来描述，如地理实体的空间幅度（Spatial Extent）这一地理特征，该元数据标准通过东西南北4个边界经度或纬度来描述该地理实体的绝对位置。

（2）数字地理空间元数据内容标准。数字地理空间元数据内容标准（最新版本为1998年版）是美国联邦地理数据委员会（Federal Geographic Data Committee，FGDC）制定的用于描述数字化的地理空间数据集的元数据标准。该标准描述了与数字化地理空间数据集定位、获取、使用和发布有关的元数据，由数据元素和复合元素的层级结构组织，定义了记录数字地理空间数据集的元数据的信息内容，同时包含元素的定义和域值。数据元素（data element）是指数据的逻辑简单项，复合元素（compound element）是一组数据元素或/和其他复合元素的组合[15]，即复合元素有下属子元素。

该元数据标准将地理数据的信息分为11个模块进行描述，并在每个模块下设具体字段以对数据或数据集进行详细描述见表4。每个模块以复合元素的名称和定义开头，之后是生产规则，即定义这一复合元素的构成。用于地理特征描述的模块为空间数据组织信息与空间参考信息两个模块。空间数据组织信息设有对数字化地理空间数据进行描述的具体字段，分为间接空间参考、直接空间参考方法、点和向量对象信息、网格对象信息4个部分，通过数据集引用地理位置的方式，以及地理位置的点、向量、网格等数据来对地理数据的位置、大小、距离等特征进行描述。空间参考信息是指数据集参考框架、编码方式、坐标描述，是空间数据组织中具体字段的参照系统。

综上，ISO 19115－2图像和栅格数据元数据标准与CSDGM元数据标准均用于描述Data.gov中的地理空间数据集。前者为国际通用元数据标准，后者为美国制定的元数据标准。二者在元数据框架构建方面存在一定的相似性，在模块设置上类似，如均包含元数据根信息、识别信息、数据质量信息、空间表示信息、空间参考信息、发布信息等，均对地理空间数据的地理特征及数据集内容、质量、创建、获取等属性进行了有效描述。

1.1.3　数据集内容描述元数据标准映射　为了便于机构各自创建数据集的POD v1.1元数据记录，FGDC成员机构构建了CSDGM和ISO 19115两类地理空间元数据标准与POD v1.1的映射[16]。ISO 19115与POD v1.1的共同元素要多于CSDGM与POD v1.1的共同元素。ISO 19115、CSDGM与POD v1.1的共同元素主要集中在POD v1.1的数据集字段与发布字段。数据集字段中三种元数据标准的共同元素有标题、描述、关键词、修订、出版者姓名、联系姓名、联系邮件、标

表 4 CSDGM 元数据标准具体元素

元数据内容	描述内容	一级元素
元数据根信息	数据内容、质量、条件及其他特征信息	识别信息、数据质量信息、空间数据组织信息、空间参考信息、实体和属性信息、发布信息、元数据参考信息
识别信息	数据集基本信息	引用、描述、内容时段、数据状态、空间域、关键词、获取限制、使用限制、联系点、浏览图片、数据集信用、安全信息、本地数据集环境、交叉参考
数据质量信息	数据集质量通用评价	属性准确性、逻辑一致性报告、完整性报告、位置准确性、所属类别、云量
空间数据组织信息	数据集的空间表示机制	间接空间参考、直接空间参考方法、点和向量对象信息|网格对象信息
空间参考信息	数据集参考框架、编码方式、坐标描述	经度坐标系统定义、纬度坐标系统定义
实体和属性信息	数据集信息内容详情,包括实体类型、属性、属性值	详细描述(实体类型、属性)、概述描述
发布信息	获取数据集的发布者和选择信息	发布者、资源描述、发布责任、标准顺序处理、定制数据处理、技术条件、可获取的时间段
元数据参考信息	元数据信息的准确性和责任主体	元数据日期、元数据编修日期、元数据未来编修日期、元数据联系、元数据标准名称、元数据标准版本、元数据时间惯例、元数据获取限制、元数据使用限制、元数据安全信息、元数据扩展
引用信息	数据集引用方式	创始人、发布日期、发布时间、标题、版本、地理空间数据表示形式、系列信息、发布信息、其他引用详情、在线连接、较大工作引用
时间信息	事件的日期和时间信息	单个日期/时间、多个日期/时间、日期/时间范围
联系信息	与数据集相关的个人和组织的身份及沟通方式	主要联系人、主要联系机构、联系位置、联系地址、联系电话、听力受损人士专用联系电话、传真电话、电子邮箱地址、服务时间、联系说明

识符、访问级别、局代码、项目代码、空间、时间、主题。发布字段中三种元数据标准的共同元素为下载 URL 与媒体类型。对于相同元素,三种元数据标准在描述方面大致相同,CSDGM、ISO 19115 在字段设置方面比 POD v1.1 更为细致具体。相同字段所对应的值有略微差别,如“修订”这一字段,POD v1.1 将其值定义为“最新更新时间(Last Update)”;CSDGM 中的对应项为“出版日期(publication date)”;ISO 19115 的对应项则包含资源维护频率、数据引用修订日期、数据第一次引用日期。尽管存在差异,三种元数据标准可以基本对应。

1.2 数据集格式描述元数据标准

数据集格式描述元数据标准是指呈现数据集内容时所采用的元数据标准,以保证数据集不仅用户可读,而且机器可读。Data.gov 使用的数据集格式描述元数据标准主要包含两种,分别为 JSON 与 ISO 19139 地理信息-元数据-XML 标准实现(ISO 19139 Geographic Information-Metadata-XML schema implementation,以下简称 ISO 19139)。

1.2.1 JSON 开放数据政策要求数据集的元数据必须以 JSON 格式的形式进行描述,以便数据目录对数据集进行统一的元数据收割。JSON 是一种易于阅读、解析、生成的轻量级文本型数据交换格式,用于优化数据交换。该格式基于两种结构构件:一是名称值对,通常通过对象、记录、结构、字典、哈希表、键列表或关联数组的形式实现;二是值的有序列表,通常通过数组、向量、列表或者序列的形式实现[17]。JSON 结构实例片段[18]如下:

```
{
  "conformsTo": "https://project-open-data.cio.gov/v1.1/schema", //表明数据集所遵循的元数据标准
  "dataset": [
    {
      "accessLevel": "public", //表明数据集的访问级别为"公开获取"
      "bureauCode": [
        "018:10" //表明数据集所属局代码为"018:10"
      ],
      "contactPoint": {
        "fn": "Jane Doe", //表明联系人
        "hasEmail": "mailto:jane.doe@agency.gov" //表明联系人邮箱
      },
      "description": "This dataset provides national statistics on the production of widgets", //对数据集内容进行概要描述
      ……
}
```

该片段对数据集所遵循的元数据标准、数据集的访问级别、局代码、联系方式、描述等字段进行了描述。由该片段可以清楚地看到数据集元数据的机器表示方式,是以“属性”:“值”的方式进行元数据描述。

1.2.2 ISO 19139 地理信息-元数据-XML 标准实现标准 ISO 19139 是 ISO 19115 地理-元数据标准(ISO 19115 Geographic Information-Metadata)的 XML 记录格式和检验规范,即 ISO 19115 的 XML 编码[19],于 2007 年发布。Data.gov 中的原始数据在内容上采用 ISO

19115－2 元数据标准，在格式上采用 ISO 19139 的 XML 记录格式来进行描述。某一数据集的 XML 结构[20]如下：

```
<gmi:MI_Metadata xmlns:gco = "http://www.isotc211.org/2005/gco" xmlns:gmd = "http://www.isotc211.org/2005/gmd" xmlns:gmi = "http://www.isotc211.org/2005/gmi" xmlns:gmx = "http://www.isotc211.org/2005/gmx" xmlns:gsr = "http://www.isotc211.org/2005/gsr" xmlns:gss = "http://www.isotc211.org/2005/gss" xmlns:gts = "http://www.isotc211.org/2005/gts" xmlns:gml = "http://www.opengis.net/gml/3.2" xmlns:xlink = "http://www.w3.org/1999/xlink" xmlns:xsi = "http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation = "http://www.isotc211.org/2005/gmi http://www.ngdc.noaa.gov/metadata/published/xsd/schema.xsd" > //命名空间声明
  <gmd:fileIdentifier > ... </gmd:fileIdentifier > //文件标识符
  <gmd:language > ... </gmd:language > //语言
  <gmd:characterSet > ... </gmd:characterSet > //字符串
  <gmd:hierarchyLevel > ... </gmd:hierarchyLevel > //所属层级
  <gmd:contact > ... </gmd:contact > //联系信息
  <gmd:dateStamp > ... </gmd:dateStamp > //日期
  <gmd:metadataStandardName > ... </gmd:metadataStandardName > //元数据标准名称
  <gmd:metadataStandardVersion > ... </gmd:metadataStandardVersion > //元数据标准版本
  <gmd:identificationInfo > ... </gmd:identificationInfo > //识别信息
  <gmd:contentInfo > ... </gmd:contentInfo > //内容信息
  <gmd:distributionInfo > ... </gmd:distributionInfo > //发布信息
  <gmd:metadataMaintenance > ... </gmd:metadataMaintenance > //元数据维护信息
</gmi:MI_Metadata >
```

上述结构展示的是数据集 XML 表示中的一级层级结构，ISO 19139 首先对描述过程中使用到的命名空间进行声明，而后逐一描述数据集的元数据信息，依次为文件标识符、语言、字符集、所属层级、联系信息、日期、元数据标准名称、元数据标准版本、识别信息、内容信息、发布信息、维护信息等。每个一级元素下设有二级、三级等描述信息。该描述层级依 ISO 19115－2 的描述层级而定。

1.3 美国开放政府数据元数据标准特点

经上述分析，美国开放政府数据元数据标准具有如下特点：一是，元数据描述详细且针对性强，该平台对数据集的元数据描述包含了数据集内容与格式两个方面，并根据平台中的数据集类型即原始数据集与地理空间数据集采用了能够凸显各自特征的不同元数据标准进行描述；二是，在现有国际通用元数据标准基础上做了本土化调整，POD v1.1 即是在 DCAT 基础上构建的元数据标准，增加了如“bureauCode”这样仅适用于美国联邦政府的字段；三 是，元数据标准与开放政府数据平台具有相同的层级结构，结构清晰，均按照数据目录－数据集－数据集发布这样的层级结构组织并描述数据集。

美国、英国、澳大利亚 3 个国家的元数据标准均包含数据目录、原始数据集、地理空间数据集 3 个方面，不同之处在于各个国家在对这三类资源进行描述时选用了不同的元数据标准。对数据目录的描述，三者均采用 DCAT 或基于 DCAT 构建的元数据标准，如美国采用 POD v1.1 元数据标准中的数据目录字段，英国基于 CKAN 记录格式，澳大利亚遵循 DCAT 元数据标准；对于原始数据集的描述，美国采用 POD v1.1 中的数据集、数据集发布字段与 JSON 分别描述其内容与格式，英国采用 CSV 和 JSON 格式，澳大利亚采用 AGLS 元数据标准；对于地理空间数据集，美国采用 ISO 19115－2、CSDGM 与 ISO 19139 描述其内容与格式，英国采用 GEMINI 地理空间元数据格式，澳大利亚则采用 ANZLIC 地理空间元数据标准。

2 对我国开放政府数据平台元数据标准构建的启示

我国目前正在进行开放政府数据平台的构建，包括国家级、地区级等不同层级的开放政府数据平台，并预计在 2018 年底前建成政府数据统一开放平台，实现公共数据资源合理适度地向社会开放[21]。而元数据标准的构建是我国开放政府数据平台构建的重要版块之一。根据 2017 年我国首个《中国地方政府数据开放平台报告》可知，我国各个地方开放政府数据平台中的元数据条目的数量及数据项并不统一，导致各个平台数据集的描述详略与呈现方式存在差异，从而对不同数据集间的整合带来障碍。黄如花与王春迎[22]在对我国已建设的 13 个地方性政府数据开放平台进行调查与分析后发现，除北京、上海、无锡、青岛的数据开放平台提供的元数据较为详细，其余均较为简单。晴青与赵荣[23]在对北京市政府数据开放现状的研究中发现，网站中的数据集大部分只提供 CSV 格式的下载，

大部分数据内容仅包含简单信息，缺乏详尽内容。陈红玉等[24]指出我国现有地方政府开放数据门户网站普遍缺乏规范的元数据标准体系，使得数据集的各项信息，包括溯源信息，很难被用户了解与利用。而赵龙文与莫荔媛等[25]在分析政府数据开放特点下的描述要求的基础上，引入DC、VoID、DCAT等元数据标准对数据资源进行目录描述、数据集、关联描述和访问描述，为开放后的数据共享、查找、管理等提供有效支持。因而，十分有必要建立统一的元数据标准，以对数据集进行统一规范的描述，并利于采用元数据收割的方式集中并及时更新各个机构或部门存储的数据集。结合美国开放政府数据元数据标准，我国在构建自身元数据标准时需注意以下几点：

2.1 元数据标准的选择应尽量广泛通用

构建统一的政府数据开放平台需要有统一的元数据标准，以确保数据集在从采集到发布的各个环节均能按照统一的规范进行。统一元数据标准的选择有两种方式：一是采用现有国际通用的开放政府数据元数据标准，如美国所采用的POD v1.1、ISO 19115－2、CSDGM、JSON等元数据标准，采用现有标准有助于实现我国数据集与其他国家数据集间的后续兼容；二是在现有国际通用元数据标准基础上，结合我国数据集的特征，制定符合我国国情的元数据标准。后者更为有效，因美国政府数据集与我国政府数据集在资源类型、数据提供单位等方面有差别。而且，我国各个地市级的政府数据开放平台均构建了各自的数据集描述方式。统一元数据标准应具有对已有地市级元数据标准的兼容性。具体构建时，可结合现有国际通用元数据标准与我国地方元数据标准，提炼出基本元数据元素与拓展元数据元素，二者分别为数据集描述中的必要字段与非必要字段，以分别应对不同详略程度的数据集描述。

2.2 元数据标准应能区分数据集类型

美国开放政府数据平台将数据集分为原始数据集与地理空间数据集两种类型，并分别采用不同的元数据标准进行描述。尤其是对于地理空间数据集，选择能够表征数据集地理特征的元数据标准进行描述。我国在构建自身元数据标准时，也应结合不同类型数据集特征，选择并制定具有针对性的元数据标准，以区分数据集类型，充分展示数据集特征，进而使数据集得到更加合理有效地使用。

2.3 元数据标准间应具有一定的互操作性

元数据标准的互操作性体现在两点：一是开放政府数据平台中的各个元数据标准间要具有互操作性，如美国开放政府数据平台所使用的元数据标准ISO 19115、CSDGM均与POD v1.1建立了映射；二是构建的元数据标准应与其参照标准具有互操作性，如POD v1.1与其构建所依据的元数据标准DCAT建立了映射，同时还与Schema.org有映射关系。因此，我国统一开放政府数据平台中的统一元数据标准既要与各地市级元数据标准建立互操作，也要与其所参照的元数据标准建立互操作。

2.4 元数据标准应同时包含数据集内容与格式的描述

数据集不仅要在内容描述上规范，同时也要在格式描述上规范，以同时实现数据集的用户可读与机器可读。数据集在内容描述时，可结合国外数据集描述标准与我国数据集特点构建符合自身数据集要求的元数据标准；而在数据集格式描述时，则可完全借鉴国外的数据集格式描述标准如JSON等。

基于上述关键点，我国开放政府数据平台元数据标准的构建应遵循以下思路。首先，确定平台中数据集的类型，对不同类型数据集做针对性的元数据描述，如参照美国、英国、澳大利亚，分别针对数据目录、原始数据集、地理空间数据集这三类主要资源选择或构建不同元数据标准；其次，确定元数据描述的基本方面，如数据内容与数据格式，在确定数据内容元数据标准的基础上，选择相应的数据格式元数据标准；再者，确定我国数据集元素属性及属性粒度，依此在DCAT、DC等现有通用元数据标准基础上，根据我国数据集特点，做本土化处理，并尽量细化描述粒度，包含从数据集来源到数据集发布过程中所涉及到的各类人、时间、机构及数据集本身的内容与格式特征，以便为用户多样化检索和数据二次开发奠定基础。

3 结语

本文以美国开放政府数据平台Data.gov为例，分析了该平台中的数据集构成及不同类型数据集的元数据标准。Data.gov平台将数据集分为原始数据集、地理空间数据集、数据工具3种类型，原始数据集采用POD v1.1元数据标准进行描述，而地理空间数据集采用ISO 19115－2与CSDGM这两种地理空间元数据标准进行描述。这3种元数据标准均是对数据集的内容信息进行描述，3种标准具有部分共有字段。除对数据集内容进行描述的元数据标准之外，还有对数据集格式进行描述的元数据标准，分别为JSON与ISO 19139元数据标准。基于对美国开放政府数据平台元

数据标准的分析，笔者认为我国在构建自身元数据标准时，需选择广泛通用的元数据标准，并对不同类型数据集采取不同的元数据标准进行描述，各个元数据标准之间要具有互操作性，并且要同时包含数据集内容与格式的元数据描述。

参考文献：

[1] Open government data [EB/OL]. [2017－03－09]. https://opengovernmentdata.org/.

[2] 郑磊，高丰. 中国开放政府数据平台研究：框架、现状与建议[J]. 电子政务，2015(7):8－16.

[3] Open data institute.. Open data certificate [EB/OL] [2017－11－17]. https://certificates.theodi.org/en.

[4] 孙璐，李广建. 政府开放数据应用分析模型构建研究[J]. 图书情报工作，2017,61(3):97－108.

[5] 于梦月，翟军，林岩. 我国地方政府开放数据的核心元数据研究[J]. 情报杂志，2016，35(12):98－104.

[6] 翟军，于梦月，林岩. 世界主要政府开放数据元数据方案比较与启示[J]. 图书与情报，2017(4):113－121.

[7] 赵蓉英，梁志森，段培培. 英国政府数据开放共享的元数据标准——对 Data.gov.uk 的调研与启示[J]. 图书情报工作，2016，60(19):31－39.

[8] 黄如花，李楠. 澳大利亚开放政府数据的元数据标准——对 Data.gov.au 的调研与启示[J]. 图书馆杂志，2017(5):87－97.

[9] 武琳，黄颖茹. 开放政府数据平台元数据标准研究进展[J]. 图书馆学研究，2017(6):14－21.

[10] 王立清，唐宇萍. 澳大利亚新西兰政府网站建设的元数据标准[J]. 情报资料工作，2004(s1):410－413.

[11] 黄如花，林焱. 国外开放政府数据描述规范的调查与分析[J]. 图书情报工作，2017,61(20):37－52.

[12] Project open data. Project open data metadata schema v1.1 [EB/OL]. [2017－02－20]. https://project-open-data.cio.gov/v1.1/schema/.

[13] W3C. Data catalog vocabulary (DCAT) [EB/OL]. [2017－02－20]. https://www.w3.org/TR/vocab-dcat/.

[14] Federal geographic data committee. Geospatial metadata [EB/OL]. [2017－03－08]. https://www.fgdc.gov/metadata.

[15] Federal geographic data committee. Organization of the standard [EB/OL]. [2017－03－09]. https://www.fgdc.gov/metadata/csdgm/organization.html.

[16] Project open data. Metadata resources for schema v1.1[EB/OL]. [2017－05－09]. https://project-open-data.cio.gov/v1.1/metadata-resources/#field-mapping.

[17] Introducing JSON [EB/OL]. [2017－05－08]. http://www.json.org/.

[18] Catalog－sample [EB/OL]. [2017－05－11]. https://project-open-data.cio.gov/v1.1/examples/catalog-sample.json.

[19] Federal geographic data committee ISO 191＊＊ suite of geospatial metadata standards [EB/OL]. [2017－03－08]. https://www.fgdc.gov/metadata/iso-suite-of-geospatial-metadata-standards.

[20] XML file [EB/OL]. [2017－05－11]. https://catalog.data.gov/harvest/object/cd79c269-d065-4a46-8a5c-0aacd2654bdb.

[21] 赵蕊菡. 政府类开放关联数据集调查研究[J]. 图书与情报，2016(4):102－112.

[22] 黄如花，王春迎. 我国政府数据开放平台现状调查与分析[J]. 情报理论与实践，2016，39(7):50－55.

[23] 晴青，赵荣. 北京市政府数据开放现状研究[J]. 情报杂志，2016，35(4):177－182.

[24] 陈红玉，翟军，袁长峰，等. 开放政府数据的溯源元数据研究及应用[J]. 情报杂志，2017，36(6):148－155.

[25] 赵龙文，莫荔媛，陈明艳. 面向政府数据开放的资源描述方法[J]. 图书情报工作,2017，61(6):115－121.

作者贡献说明：

司莉：指导论文撰写，论文思路修正；

赵洁：确定选题，网站调研与文献查阅，撰写与修改论文。

Investigation and Enlightenment of Metadata Standards of American Open Government Data

Si Li[1] Zhao Jie[2]

[1] Center for Studies of Information Resources, Wuhan University, Wuhan 430072

[2] School of Information Management, Wuhan University, Wuhan 430072

Abstract: [**Purpose/significance**] This paper takes the metadata standards of Data.gov, an open government data website, as an example, and analyzes its metadata system and specific standards so as to provide references for the construction of our country's open data metadata standards. [**Method/process**] Using the method of the case analysis, the paper summarized the system structure of the metadata standard of American open government data. [**Result/conclusion**] The metadata standards of American open government data can be divided into two categories, which are dataset content metadata standards and dataset format metadata standards. Data.gov uses different metadata standards to describe original datasets and geospatial datasets respectively. We can reference the metadata standards system in Data.gov in the construction of metadata standards of Chinese open government data.

Keywords: metadata open government data metadata standard Data.gov